

ІДЕНТИФІКАЦІЯ СМИСЛОВИХ ВІДНОШЕНЬ У ТЕКСТАХ ВІКІПЕДІЇ ДЛЯ ПОБУДОВИ СЕМАНТИЧНОЇ МЕРЕЖІ

Петрасова С.В., Шанідзе О.Д., Швець С.І.

*Національний технічний університет
«Харківський політехнічний інститут», м. Харків*

У зв'язку зі зростаючим обсягом інформації, що потребує систематизації та обробки, ускладнюється процес отримання необхідних даних з текстів природної мови. Тому розробка засобів зберігання та швидкої і ефективної обробки даних стає актуальним завданням Natural Language Processing, зокрема Information Extraction (IE).

Відповідно до інформації, що видобувається, ІЕ включає наступні задачі: визнання іменованих сутностей, виявлення атрибутів / відношень, видобування фактів / подій.

Одним з найбільш складних завдань є отримання інформації про семантичні (смислові) відношення між об'єктами. Семантичне відношення встановлюється між лексичними одиницями в межах певного семантичного поля та може виражати відношення класифікації, гіпонімії / гіперонімії, міронімії / холонімії, синонімії, антонімії та інші.

Для ідентифікації семантичних відношень застосовуються методи, засновані на правилах, або методи машинного навчання: наївний класифікатор Баєса, дерева рішень, метод опорних векторів, приховані марковські моделі, тощо.

У роботі пропонується алгоритм автоматичного визначення семантичних відношень за допомогою підходу, що базується на правилах. Для цього було розроблено корпус текстів Вікіпедії, представлених статтями категорії «інформаційні технології».

На першому етапі виявляємо предикати, наприклад, *include*, *consist of*, *contain*, визначені як семантичні відношення. На наступному етапі ідентифікації смислових відношень відбувається пошук понять, зв'язаних попередньо виявленими предикатами, та їхніх синонімів, отриманих з WorldNet. Останнім кроком є екстракція дієслів між виявленими поняттями та їхніми синонімами, що представляють семантичні відношення.

Отже, в результаті ми отримуємо семантичну інформацію з корпусу текстів, тобто семантичну мережу слів, зв'язаних смисловими відношеннями.

Література:

1 Петрасова С.В. Формалізація інформаційно-лінгвістичних сутностей та відношень засобами штучного інтелекту. Інтелектуальні системи та прикладна лінгвістика : матеріали V Всеукр. науково-практичної конференції, 14 квітня 2016 р. Харків: НТУ «ХПІ», 2016. С. 48–49.